图书情报工作
第 65 卷 第 16 期 2021 年 8 月

社交网络用户自我披露意愿研究*
——以新浪微博为例

■ 臧国全[1,2] 孔小换[1] 张凯亮[3] 于政杰[1]
[1] 郑州大学信息管理学院 郑州 450001 [2]郑州市数据科学研究中心 郑州 450001
[3] 郑州大学政治与公共管理学院 郑州 450001

摘 要：[目的/意义]社交网络用户自我披露对以用户生成内容为业务基础的社交网络具有战略意义，而用户生成内容的数量和质量取决于用户自我披露意愿。因此，研究社交网络用户自我披露意愿及其影响因素，为社交网络平台制定隐私政策、提升用户自我披露水平提供参考，促进社交网络平台的健康快速发展。[方法/过程]参考已有研究框架，构建社交网络用户自我披露意愿的研究模型。选择新浪微博作为社交网络平台代表，采用 Python 爬虫方法获取用户微博数据，据此分析用户自我披露意愿。[结果/结论]微博内容的语义、位置标签和数据权限均影响用户自我披露意愿，隐藏位置标签和设置数据权限等操作会显著提高用户自我披露意愿。社交网络用户自我披露意愿是一种个人特质，受性别、年龄、学历等人口统计学因素影响。

关键词：社交网络 自我披露 披露意愿 新浪微博
分类号：G251
DOI：10.13266/j.issn.0252－3116.2021.16.010

1 引言

自我披露指个人与他人交往时，自愿且真实地向目标人倾诉自己的思想、展示自己的行为[1]。在现代社会中，自我披露是人类各种社会关系得以形成的前提条件，是人类传播的基础。社交网络可以帮助人们建立和维护社会关系，具有鼓励用户自我披露的天然机制，越来越多的用户开始在社交网络中进行自我披露，以微信、微博、抖音和快手等为代表的社交媒体已广泛渗透到人们的日常生活。然而，近年来层出不穷的用户隐私泄露事件(例如，“1.59 亿领英用户敏感数据遭泄露[2]”“8 700 万 Facebook 用户数据被不当泄露给政治咨询公司剑桥分析[3]”等)，已引发广泛的社会关注，如何鼓励用户持续自我披露成为社交网络平台发展的首要问题。用户是社交网络的基本要素，其快速发展依赖于用户自我披露内容的数量和质量。社交网络用户自我披露的历史数据是用户自我披露行为的真实记录，笔者选择新浪微博作为社交网络平台代表，通过爬虫技术获取用户自我披露的历史微博数据，根据微博数据特征，从微博语义内容、微博位置标签和微博数据权限 3 个维度探究社交网络用户自我披露意愿及其影响因素。研究社交网络用户的自我披露意愿有助于分析社交网络用户自我披露的影响因素，为社交网络平台制定隐私政策、提升用户自我披露水平提供参考。另外，社交网络中的内容会对用户(尤其是青少年用户)的自我披露意愿产生深刻影响，了解影响用户自我披露和隐私保护的因素，有助于引导用户(尤其是青少年用户)正确设定隐私设置和进行适当的自我披露。

2 文献综述

目前，已有大量关于用户自我披露的实证研究，笔者从数据获取方法、重要影响因素和研究情景等角度对相关研究进行归纳，见表 1。

根据表 1 可以得出，现有关于用户自我披露的研究，将用户自我披露的影响因素主要归纳为“感知风险

* 本文系国家自然科学基金项目“数字保存的风险型元数据与风险监控研究”(项目编号：71673255)研究成果之一。
作者简介：臧国全(ORCID：0000－0002－9606－6455)，院长，教授，博士生导师；孔小换(ORCID：0000－0001－9858－016X)，硕士研究生；张凯亮(ORCID：0000－0003－0454－0525)，博士研究生，通讯作者，E-mail：779015223@qq.com；于政杰(ORCID：0000－0003－2707－0669)，硕士研究生。
收稿日期：2021－03－29 修回日期：2021－06－15 本文起止页码：90－97 本文责任编辑：徐健

表 1　用户自我披露相关研究汇总

数据获取方法	用户自我披露的影响因素	研究情境	文献来源
网络问卷	感知风险、感知收益、隐私关注	社交网络	H. Krasnova 等[4]（2010）、郭海玲等[5]（2019）
		电子商务	T. Dinev 等[6]（2004）、朱侯等[7]（2017）
		在线医疗	张星等[8]（2016）、王瑜超[9]（2018）
	态度、主观规范、行为意愿、感知控制	社交网络	H. Xu 等[10]（2013）、李征仁[11]（2014）
		电子商务	朱侯等[7]（2017）、刘百灵等[12]（2017）
		在线医疗	李裕广[13]（2015）、张星等[8]（2016）
	感知控制、信任、隐私规则	社交网络	郭海玲等[5]（2019）、N. Zlatolas 等[14]（2019）
		电子商务	刘百灵等[12]（2020）
	信任、感知风险、感知收益	社交网络	H. Ko 等[15]（2009）、E. T. Loiacono[16]（2015）
		在线医疗	姜又琦[17]（2017）、王瑜超[9]（2018）
实验调查	信任、隐私关注、感知收益	电子商务	J. C. Zimmer 等[18]（2010）、K. Premazzi 等[19]（2014）
	感知收益、感知风险、态度、社会规范、感知控制	社交网络	T. Stern 等[20]（2015）、E. Kim 等[21]（2016）
	信任、隐私关注、感知风险	在线医疗	G. Bansal 等[22]（2010）

和感知收益（隐私计算理论）”“行为态度、主观规范和感知控制（计划行为理论）”“隐私规则、信任和感知控制（沟通隐私管理理论）”等。相关研究大多以隐私计算理论为基础，结合计划行为理论和沟通隐私管理理论等，根据不同的研究情境，通过增删变量的方式构建不同的用户自我披露研究模型。而上述影响用户自我披露的因素大多是用户的主观感知因素，因此，数据调查方式也多为问卷调查，然而该方法具有一定局限性：①问卷题项设计固定，缺乏弹性，揭示现象的角度有限；②问题设计可能存在倾向性，容易对被试者产生引导作用；③被试者自我报告的数据主观性较强，不一定真实反映其态度及行为。因此，部分研究将实验法和问卷调查法相结合，先在模拟情景中观测用户的披露意愿或实际披露行为，随后让用户填写影响因素的量表问题以获取相关数据，但该方法在问卷设计及填写方面仍然具有较强主观性，且实验成本较高。

笔者为克服上述研究方法存在的客观缺憾，采用爬虫技术抓取用户在社交网络上自我披露的历史数据，挖掘社交网络用户的自我披露行为特征。该方法的优点有 2 个：①社交网络平台保留了大量用户自我披露的历史数据，是用户自我披露的真实行为记录。因此，可对用户自我披露的历史数据进行深度挖掘，以分析用户的行为特征。②采用爬虫技术获取的用户数据体量大，且真实可靠，能在很大程度上克服问卷调查法的主观性过强的弊端，更客观反映用户行为特征。

鉴于此，笔者采用爬虫技术，抓取用户在社交网络上自我披露的历史数据，基于 A. Adams 等[23]提出的隐私感知模型，从语义维度、空间维度（包含空间类型和空间级别）和数据权限维度分析用户的自我披露意愿及其影响因素，基于数据分析结果向社交网络平台提出针对性的运营建议。

3　用户自我披露意愿研究模型

A. Adams 等采用扎根理论方法对已有相关研究进行归纳，从信息敏感性、信息接收和信息使用 3 个维度构建用户隐私感知模型，逐渐成为研究用户隐私问题的有效框架；J. S. Cunningham 等[24]基于该模型，针对数字照片在线收藏与分享情境，从信息敏感性、信息接收者和信息使用 3 个维度研究了用户对个人数字照片共享意愿的影响因素；C. Jennett 等[25]利用该模型，针对信用贷款申请情景，从数据敏感性、数据接收者和数据使用 3 个维度研究了贷款申请人向贷款提供者披露个人信息意愿的影响因素。

用户在社交网络中发布信息的过程就是用户的自我披露过程，因此，笔者借鉴 A. Adams 的隐私感知模型，分析用户在社交网络中的自我披露意愿。然而，在社交网络情景下，用户自我披露的大多数内容面向所有用户，无法预测信息使用者。因此，笔者仅选择 A. Adams 隐私模型中的“信息敏感性”和“信息接收”两个维度研究社交网络用户自我披露行为。社交网络情景下，“信息敏感性”指用户自我披露内容敏感性，包含语义、空间两个维度；“信息接收”是指用户对其自我披露内容设置的数据权限。

3.1　语义维度

语义维度是指基于用户在社交网络中发布信息的语义内容特征，从语义层面将用户自我披露的文本内容划分为若干维度。目前，已有许多关于信息类型对用户行为影响的研究结论。比如，信息类型对社交网

络用户的转发意愿有显著影响，用户转发享乐型信息的意愿显著高于功利性信息[26]；在热点事件的微博话题讨论中，用户对非情绪化信息的分享意愿高于情绪化信息，对积极情绪信息的分享意愿高于消极情绪信息[27]。

笔者参考侯敏[28]对微博信息的分类结果以及用户发布信息的目的，从信息的语义内容角度，将社交网络用户自我披露的语义内容划分为“生活记录类”“新闻娱乐类”“工作学习类”和“情感表达类”，以分析不同语义维度下用户自我披露意愿的差异。

3.2 空间维度

空间维度指基于用户在社交网络上发布信息时的位置标签内容特征，从空间级别和空间类型层面将用户自我披露的位置标签划分为若干维度，包含空间级别和空间类型两个子维度。

空间级别子维度。社交网络平台允许用户在发布信息时自定义位置标签级别，可以是当前所在的城市、街道，也可精确到某一具体地点，用户可通过空间级别的个性化设置保护其个人隐私。因此，笔者将空间级别设置为“城市”“街道”“具体位置”3 个水平。

空间类型子维度。只有当空间级别为“具体位置”水平时，才可分析空间类型对用户自我披露意愿的影响。已有研究表明[29-30]，用户对不同地理位置类型的敏感性存在显著差异，大众娱乐场所是敏感性较弱的地点，与身体健康相关的地点敏感性较高，而居住地被认为是敏感性最强的地点。另外，根据《中国互联网络发展状况统计调查》[31]，学生占我国网民群体的比例最高（占比 21%），而学生的大部分活动地点为学校。因此，笔者将空间类型设置为“大众娱乐场所”“医疗健康场所”“学习教育场所”和“个人居住场所”4 种空间类型。

3.3 数据权限维度

数据权限是指用户对数据接收者范围的控制，用于研究不同范围的数据接收者对用户自我披露意愿的影响。研究表明[32-33]，赋予用户对数据接收者范围的选择权能有效降低用户的隐私担忧，从而提高用户自我披露的意愿。大多数社交网络平台均允许用户对其自我披露内容进行数据权限设置。例如，新浪微博内容的数据权限分为：公开可见、粉丝可见、好友圈可见、仅自己可见等；Facebook 的数据权限分为：仅自己可见、朋友可见、朋友的朋友可见、所有人可见等。为便于后续研究，笔者将数据权限简化为“所有人可见”和“部分人可见”两个水平。

基于上述维度分析，构建社交网络用户自我披露意愿的研究模型，如图 1 所示：

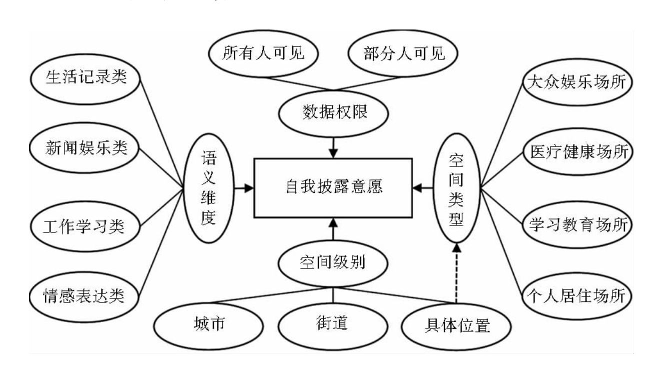


图 1 社交网络用户自我披露意愿的研究模型

4 数据获取与处理

根据新浪官方统计数据[34]，截至 2020 年 9 月，新浪微博月活跃用户已达 5.11 亿，平均日活跃用户为 2.24亿，新浪微博已成为国内用户网络社交的重要媒介。因此，笔者选择新浪微博作为社交网络平台代表，研究社交网络用户自我披露的意愿。

4.1 获取方法及数据项

获取用户数据的方法主要有两种：①通过新浪微博 API 接口获取；②通过网络爬虫获取。前者需获得新浪微博授权，且请求次数有限制。因此，笔者采用网络爬虫方法获取微博用户数据。

根据上述构建的社交网络用户自我披露意愿的研究模型，需获取的数据项包括两部分：①用户的基本数据，包括性别、出生日期、教育背景等人口统计数据以及用户注册时间；②用户的历史微博数据，包含用户已发布微博的正文文本、发布微博时的位置标签、已发布的微博总数等。

4.2 用户样本选择

为保证用户样本的真实性和合理性，笔者从账户活跃度、账户实际运营者、用户整体特征 3 个方面对用户样本进行筛选和控制。

（1）账户活跃度，应选择活跃度相对高的账户。目前缺乏这类账户的界定标准，笔者设定的筛选标准为：账户注册时间满 1 年以上，且发布微博总数不低于 20 条。

（2）账户实际运营者，应是对信息披露行为具有主导权的普通个人用户，而非官方微博、明星、专业博主等机构运营账户。为简化筛选条件，剔除粉丝数量大于 5 000 的账户。

（3）用户整体特征，应符合整体微博用户特征。

臧国全，孔小换，张凯亮，等.社交网络用户自我披露意愿研究——以新浪微博为例[J].图书情报工作,2021,65(16):90－97.

将用户样本的性别、年龄和学历层次与《2018 年微博用户发展报告》[35] 中的整体微博用户特征进行对比，二者相差不大即可，否则应调整用户样本。

4.3 数据处理

4.3.1 数据清洗

笔者选择的数据清洗工具为 Pandas 工具包(Python 的第三方库)，包括用户基本数据项清洗和微博文本数据清洗。

基本数据项清洗：剔除不完整、逻辑错误和重复的数据。不完整数据是指用户基本数据项缺失的数据，剔除不完整数据以便于后续人口统计特征分析。逻辑错误数据是指用户年龄与注册时间、年龄与教育背景等存在明显逻辑错误的数据。比如，“出生日期”应早于“注册时间”，且两者差值不低于 14(微博的推荐使用年龄为 14 岁及以上)；“教育背景”为“高中”的用户年龄应不低于 15 周岁，“教育背景”为“大学”的用户年龄应不低于 17 周岁；等。重复数据是指用户 ID 重复的数据，采用“滚雪球”的方式会存在多次抓取同一用户的情况，应剔除重复用户。

微博文本数据清洗：原始微博文本为自然语言，需对其进行过滤和清洗，以便进行语义分类。主要包括：①删除非原创微博内容；②删除“@ 用户名”；③删除微博正文中的图片、视频、网页等 URL 链接。

4.3.2 数据分类

(1)语义分类。语义分类是指将用户发布的微博按照语义内容划分为若干类别，以分析不同语义类型对用户自我披露意愿的影响。首先，利用 python 的第三方库“jieba”分词和去除停用词，将微博转换为纯词组的文本，并将所有词组整合形成一个纯词组的集合；其次，使用词权重计算算法(Term Frequency - Inverse Document Frequency，TF-IDF)计算单个词组在整个词组集合中的权重，选择权重较高的词组作为文本聚类模型的特征词，从而实现微博文本向量化；再次，使用主题挖掘模型(Latent Dirichlet Allocation，LDA)进行文本主题挖掘，形成若干聚类簇；最后，基于聚类簇中的高权主题词，将各聚类簇分别归类入生活记录类、新闻娱乐类、工作学习类和情感表达类 4 种类型中，实现语义分类目的，示例见表 2。

表 2 LDA 主题挖掘结果及高权词

类别	LDA 主题	高权词(权重)
生活	0#	旅行(0.017)、在家(0.012)、放假(0.011)
记录类	6#	希望(0.018)、努力(0.017)、生活(0.014)、健康(0.013)
	9#	红包(0.07)、感谢(0.033)、睡(0.025)、拍照(0.014)
	11#	吃(0.096)、买(0.025)、晚上(0.016)、东西(0.012)、每天(0.011)
新闻	2#	视频(0.14)、榜(0.066)、明星(0.018)
娱乐类	3#	打榜(0.03)、新歌(0.027)、音乐(0.023)
	7#	文章(0.025)、头条(0.022)、信息(0.016)
	8#	中国(0.03)、活动(0.023)、发布(0.022)、电影(0.017)
	10#	病例(0.148)、病毒(0.043)
工作	4#	学院(0.013)、老师(0.012)、学生(0.007)
学习类	12#	教师(0.059)、爱心(0.037)、教育(0.011)
	13#	能量(0.026)、学校(0.023)、计划(0.012)、开学(0.011)
	14#	守护者(0.036)、加入(0.022)、青春(0.013)、志愿者(0.006)
情感	1#	开心(0.04)、感觉(0.03)、欢乐(0.003)
表达类	5#	哈哈哈(0.056)、快乐(0.044)、早安(0.025)、喜欢(0.008)
	15#	死亡(0.023)、觉得(0.017)、喜欢(0.013)、感觉(0.008)

(2)空间分类。空间分类是指对用户发布微博时选择的位置标签进行分类，包括空间级别分类和空间类型分类，以分析微博位置标签的空间级别和空间类型对用户自我披露意愿的影响。

空间级别分类。首先，根据《2020 年中华人民共和国行政区划代码》[36] 将位置标签代码化。其次，采用条件判断语句对位置标签代码进行判断。若位置标签代码为空，则将该标签归入“无位置标签”类；若位置标签代码非空，当位置标签代码最后两位为“00”时，将该标签的空间级别设置为“城市”类；当位置标签代码最后两位非“00”时，检验位置标签文本，若位置标签文本的最后一个字为“道”或“街”或“路”或“区”或“巷”或“弄”或“里”，将该标签的空间级别设置为“街道”类。最后，将不符合上述条件的位置标签设置为“具体位置”类。

空间类型分类。当位置标签的空间级别为“具体位置”类时，讨论该标签的空间类型才具有实际意义。笔者参考相关研究，将这部分微博位置标签的空间类型划分为大众娱乐场所、医疗健康场所、学习教育场所和个人居住场所 4 类，采用模糊匹配算法对位置标签的空间类型进行标记分类。比如，将商超、饭店、酒店、景区等归为大众娱乐场所；将医院、门诊部、诊所、疗养院、养老院等归为医疗健康场所；将大中小学校、科研院所等归为学校教育场所；将住宅区归为个人居住场所。

5 数据统计与分析

5.1 数据统计

2020 年 12 月 10 日至 2021 年 01 月 31 日期间，采用自编 Python 爬虫程序，以某一用户的粉丝列表为起

点，采用“滚雪球”的方式共获取 61 902 个用户 ID，经用户样本条件筛选和数据清洗后，剩余 2 047 个有效用户 ID，包含约 15.7 万条微博。用户样本的人口统计学特征如表 3 所示：

表 3 用户样本的人口统计学特征

特征	类别	频次	频率/%
性别	男	923	45.1
	女	1 124	54.9
年龄	19 岁及以下	427	20.9
	20－29 岁	574	28.0
	30－39 岁	552	27.0
	40 岁及以上	494	24.1
学历	高中及以下	580	28.3
	专科	876	42.8
	本科及以上	591	28.9

5.2 数据分析

用户发布微博时设置的数据权限能反映其自我披露的意愿，当用户将数据权限设置为“所有人可见”时的自我披露意愿高于“部分人可见”，正是由于用户对其自我披露内容的数据权限设置，爬虫软件只能爬取到的数据权限为“所有人可见”的微博。因此，统计爬取到的用户微博数量和该用户发布的微博总数，二者的比例在一定程度上能够反映该用户的自我披露意愿，即“公开发布微博数量比例”代表用户自我披露意愿。笔者采用 SPSS 22.0 分别分析人口统计学因素、语义维度、空间维度和数据权限对用户自我披露意愿的影响。

5.2.1 人口统计学因素与用户自我披露意愿

自变量为用户性别、年龄、学历等人口统计学变量，且均为虚拟变量。因此，性别变量中：“男”＝1，“女”＝2；年龄变量中：“19 岁及以下”＝1，“20－29 岁”＝2，“30－39 岁”＝3，“40 岁及以上”＝4；学历变量中：“高中及以下”＝1，“专科”＝2，“本科及以上”＝3。因变量为用户自我披露意愿。对其进行多元线性回归，结果如表 4 所示：

表 4 多元线性回归结果

模型	非标准化系数		标准化系数	t	Sig
	B	标准误差	Beta		
（常量）	0.746	0.033		22.685	0.000
性别	－0.040	0.014	－0.063	－2.869	0.004
年龄	－0.014	0.006	－0.046	－2.105	0.035
学历	－0.026	0.009	－0.062	－2.830	0.005

性别变量。根据表 4 可以得出，用户性别显著负向影响其自我披露意愿，即女性用户自我披露意愿显著低于男性。相关研究[37]将此现象解释为：与女性相比，男性更希望通过社交网络来结识陌生人（尤其是异性），并发展建立起新的关系。因此，男性用户更倾向于在社交网络中自我披露。

年龄变量。根据表 4 可以得出，用户年龄显著负向影响其自我披露意愿，即高年龄用户自我披露意愿显著低于低年龄用户。相关研究[38]表明，年长者一般比年轻人有更高的隐私顾虑，也更可能担忧因自我披露造成的负面影响。另外，大多数年轻人将社交网络视为其展现个性、塑造身份认同的重要平台，从而在社交网络中进行更多的自我披露。因此，随着用户年龄增长，其在社交网络平台的自我披露意愿会逐渐降低。

学历变量。根据表 4 可以得出，用户学历显著负向影响其自我披露意愿，即高学历用户自我披露意愿显著低于低学历用户。相关研究表明，高学历用户的隐私关注[39]（或隐私忧虑[40]、隐私敏感性[41]）程度显著高于低学历用户，而隐私关注负向影响用户自我披露意愿，因此高学历用户在社交网络中的自我披露水平相对较低，与已有研究结论相符。

综上，用户自我披露意愿是一种受性别、年龄和学历等人口统计学特征综合影响的个人特质，本研究关于人口统计学因素对用户自我披露意愿的检验结果可作为对已有研究结论的补充证明。

5.2.2 语义维度与用户自我披露意愿

基于语义分类结果，可获取用户公开发布微博中各种语义类型微博的数量分布。经统计，用户在 4 种语义类型下公开发布微博的平均数量分别为 163、212、148 和 245。为检验同一用户在不同语义维度下公开发布的微博数量是否具备显著性差异，笔者采用配对样本 T 检验，基于语义类型两两配对，以分析不同语义类型对用户自我披露意愿的影响，检验结果如表 5 所示：

表 5 不同语义类型之间的配对样本 T 检验结果

	生活纪录类	新闻娱乐类	工作学习类	情感表达类
生活纪录类	1			
新闻娱乐类	0.003	1		
工作学习类	0.137	0.000	1	
情感表达类	0.001	0.152	0.000	1

根据表 5 可以得出，用户在“新闻娱乐类”和“情感表达类”维度下公开发布的微博数量均显著高于“生活记录类”和“工作学习类”，而“新闻娱乐类”和“情感表达类”之间以及“生活记录类”和“工作学习类”之间的微博数量差异并不显著。可能是由于不同语义类型的微博包含的个人隐私信息量存在差异，从

而导致用户隐私泄露概率不同。比如,“新闻娱乐类”微博内容大多为公开的公共信息,而“生活记录类”可能包含姓名、住址、职业等个人敏感信息。相比之下,发布“新闻娱乐类”微博导致用户隐私泄露的可能性远低于“生活记录类”微博。因此,用户发布“新闻娱乐类”微博的意愿高于“生活记录类”,这与石文华等[26]的研究结论一致,即用户更喜欢分享享乐型信息。

5.2.3 空间维度与用户自我披露意愿

(1)位置标签对用户自我披露意愿的影响。用户发布微博时可自主选择是否添加位置标签,因此,分析空间维度对用户自我披露意愿的影响时,首先应分析有位置标签的微博数量和无位置标签的微博数量的差异。经统计,用户在有位置标签和无位置标签维度下公开发布微博的平均数量分别为355和413,对其进行配对样本T检验。检验结果显示同一用户发布有位置标签的微博数量显著低于无位置标签的微博数量(P=0.026),表明用户发布微博时更倾向于隐藏位置标签。可能是由于位置信息是用户敏感度较高的个人隐私信息,将用户的历史位置数据与用户公开的其他数据进行关联分析,可能推测出用户的敏感信息甚至是建立用户画像(比如,D. Preotiuc 等[42]通过分析 Foursquare 用户的签到数据,将用户划分为“普通用户”“商人”和“工作狂”等类型,并预测了用户未来的动向)。因此,自我披露时的位置标签可能导致用户的其他隐私信息泄露,从而显著降低社交网络用户的自我披露意愿。

(2)空间级别对用户自我披露意愿的影响。基于空间级别分类结果,统计用户在不同空间级别发布微博的数量分布,结果显示用户在城市、街道和具体位置3种空间级别下公开发布微博的平均数量分别为153、111和91。采用配对样本T检验,基于空间级别两两配对,分析同一用户在不同空间级别的自我披露意愿是否具备显著性差异,检验结果如表6所示:

表6 不同空间级别之间的配对样本T检验结果

	城市	街道	具体位置
城市	1		
街道	0.016	1	
具体位置	0.000	0.030	1

根据表6可以得出,空间级别对用户自我披露意愿有显著影响,空间级别越低(即位置标签越具体),用户自我披露意愿也越低。因为位置标签越具体,泄露个人隐私信息的可能性越高。因此,位置标签的空间级别正向影响用户的自我披露意愿。

(3)空间类型对用户自我披露意愿的影响。基于空间类型分类结果,统计用户在不同空间类型发布微博的数量分布,结果显示用户在大众娱乐场所、医疗健康场所、学习教育场所和个人居住场所4种空间类型下平均发布微博数量分别为32、15、25和19。采用配对样本T检验,基于空间类型两两配对,分析同一用户在不同空间类型维度下的自我披露意愿是否具备显著性差异,检验结果如表7所示:

表7 不同空间类型之间的配对样本T检验结果

	大众娱乐场所	医疗健康场所	学习教育场所	个人居住场所
大众娱乐场所	1			
医疗健康场所	0.001	1		
学习教育场所	0.156	0.000	1	
个人居住场所	0.010	0.085	0.000	1

根据表7可以得出,用户在大众娱乐场所和学习教育场所的自我披露意愿均显著高于医疗健康场所和个人居住场所,但在大众娱乐场所和学习教育场所之间以及医疗健康场所和个人居住场所之间的自我披露意愿没有显著差异。可能是因为大众娱乐场所和学习教育场所属于公共场所,医疗健康场所较之大众娱乐场所和学习教育场所私密性更高,而个人居住场所属于私人场所,私密性最高,相关研究表明[30],用户在公共场所的自我披露意愿显著高于私人场所。另外,基于医疗健康场所的位置标签与用户的其他相关数据可能推测出用户的身体健康状况,从而导致用户的医疗健康隐私泄露,因此社交网络用户在医疗健康场所的自我披露意愿较低。

5.2.4 数据权限与用户自我披露意愿

通过爬虫获取的微博均为用户公开发布的微博(即数据权限为“所有人可见”的微博),根据用户发布的历史微博总数和用户公开发布的微博数量,可计算获得用户非公开发布的微博数量。经统计,用户发布的数据权限为“所有人可见”和“部分人可见”的微博平均数量分别为768和945,对其进行配对样本T检验,检验结果表明同一用户发布有数据权限的微博数量显著高于无数据权限的微博数量(P=0.005),即控制数据权限正向影响用户自我披露意愿。已有研究表明[43],用户在披露隐私信息时会进行数据权限控制,添加数据权限可以有效降低用户的隐私担忧。

6 结语

综上所述,笔者得出如下研究结论:①社交网络用户自我披露意愿是一种个人特质,受性别、年龄和学历等人口统计学特征影响,相关检验结果与已有研究结论一致;②自我披露信息的语义类型对用户自我披露

意愿产生显著影响，控制位置标签和数据权限会显著提高用户自我披露意愿。

本研究结果可应用于：①用户自我披露的信息类型取决于用户自身，社交网络平台无法控制，但可通过更改位置标签和数据权限的默认首选项，提高用户对其自我披露的感知控制。比如，将用户自我披露时的位置标签默认设置为“无”或将位置标签的空间级别默认设置为“城市”，也可将用户自我披露时的数据权限默认设置为“部分人可见”。②将自我披露时附带的位置标签、数据权限等选择权赋予用户，提高用户对其自我披露的感知控制能力，降低因平台默认设置而导致用户个人隐私信息泄露的可能性，从而提高用户持续使用社交网络进行自我披露的意愿。

本研究的局限性主要有：①在用户样本筛选过程中，有效用户率较低（仅 3.3%），主要是因为用户的个人基本数据不完善，即用户在社交网络中披露个人基本数据的意愿偏低；②由于篇幅限制，笔者仅从用户发布内容的角度分析其自我披露意愿，今后的研究可从用户披露个人基本数据的角度出发，分析其自我披露意愿及其影响因素；③采用“公开发布微博数量比例”代表用户自我披露意愿的准确度可能有待商榷，期望能够在未来研究中综合采用多指标衡量用户自我披露意愿。

参考文献：

[1] JOURARD S M, LASAKOW P. Some factors in self-disclosure [J]. Journal of abnormal & social psychology, 1958, 56(1):91-98.

[2] 搜狐网. LinkedIn 遭黑客攻击，约 1.59 亿用户数据被窃[EB/OL]. [2021-01-14]. https://m.sohu.com/a/292645813_557054.

[3] 百度百科. 剑桥分析丑闻[EB/OL]. [2021-01-14]. https://baike.baidu.com/item/%E5%89%91%E6%A1%A5%E5%88%86%E6%9E%90%E4%B8%91%E9%97%BB/49935337?fr=aladdin.

[4] KRASNOVA H, KOLESNIKOVA E, GÜNTHER O. Leveraging trust and privacy concerns in online social networks: an empirical study[C]// 18th European conference on information systems. Illinois: AIS elibrary, 2010.

[5] 郭海玲，马红雨，许泽辉. 社会化媒体用户信息披露意愿影响模型构建与实证－以微信用户为例[J]. 图书情报工作，2019，63(15):111-120.

[6] DINEV T, HART P. Internet privacy concerns and their antecedents-measurement validity and a regression model[J]. Behaviour & information technology, 2004, 23(6):413-422.

[7] 朱侯，王可，严芷君，等. 基于隐私计算理论的 SNS 用户隐私悖论现象研究[J]. 情报杂志，2017，36(2):134-139.

[8] 张星，陈星，侯德林. 在线健康信息披露意愿的影响因素研究：一个集成计划行为理论与隐私计算的模型[J]. 情报资料工作，2016(1):48-53.

[9] 王瑜超. 在线医疗社区用户健康隐私信息披露意愿的影响因素研究[J]. 信息资源管理学报，2018，8(1):93-103.

[10] XU H, TEO H, TAN B C Y, et al. The role of push-pull technology in privacy calculus: the case of location-based services[J]. Journal of management information systems, 2010, 26(3):135-174.

[11] 李征仁. 移动互联网环境下用户隐私关注的影响因素及隐私信息扩散规律研究[D]. 北京：北京邮电大学，2014.

[12] 刘百灵，夏惠敏，李延晖. 移动购物用户信息披露意愿影响因素的实证研究－基于公平理论和理性行为理论视角[J]. 情报理论与实践，2017，40(5):87-93.

[13] 李裕广. 在线医疗社区患者医疗信息共享意愿影响因素研究[D]. 哈尔滨：哈尔滨工业大学，2015.

[14] ZLATOLAS N, WELZER T, HÖLBL M, et al. A model of perception of privacy, trust, and self-disclosure on online social networks [J]. Entropy, 2019, 21(8): 772-789.

[15] KO H, CHEN T. Understanding the continuous self-disclosure of bloggers from the cost-benefit perspective[C]// Proceedings of the 2nd conference on human system interactions. New York: IEEE Press, 2009.

[16] LOIACONO E T. Self-disclosure behavior on social networking web sites[J]. International journal of electronic commerce, 2015, 19(2):66-94.

[17] 姜又琦. 在线医疗网站用户个人健康信息披露意愿影响因素研究[D]. 武汉：武汉大学，2017.

[18] ZIMMER J C, ARSAL R, AL-MARZOUQ M, et al. Knowing your customers: using a reciprocal relationship to enhance voluntary information disclosure[J]. Decision support systems, 2010, 48(2): 395-406.

[19] PREMAZZI K, CASTALDO S, GROSSO M, et al. Customer information sharing with e-vendors: the roles of incentives and trust [J]. International journal of electronic commerce, 2014, 14(3): 63-91.

[20] STERN T, SALB D. Examining online social network use and its effect on the use of privacy settings and profile disclosure[J]. Bulletin of science technology & society, 2015, 35(1):25-34.

[21] KIM E, LEE J, SUNG Y, et al. Predicting selfie-posting behavior on social networking sites: an extension of theory of planned behavior[J]. Computers in human behavior, 2016, 62(9):116-123.

[22] BANSAL G, ZAHEDI F M, GEFEN D. The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online[J]. Decision support systems, 2010, 49(2):138-150.

[23] ADAMS A, SASSE M A. Privacy in multimedia communications: protecting users, not just data [M]. London: Springer, 2001.

[24] CUNNINGHAM J S, MASOODIAN M, ADAMS A. Privacy issues for online personal photograph collections[J]. Journal of theoretical & applied electronic commerce research, 2010, 5(2):26-40.

[25] JENNETT C, BROSTOFF S, MALHEIROS M, et al. Adding insult to injury: consumer experiences of being denied credit[J]. International journal of consumer studies, 2012, 36(5):549-555.

[26] 石文华，陈楠，张绮. 关系类型和信息类型对社交网络用户转发意愿的影响研究[J]. 制度经济学研究，2016(4):162-180.

[27] 万晓榆，杨腾，袁野，等. 微博信息情绪类型对用户分享意愿的

影响研究[J].情报科学,2019,37(8):97－107.
［28］侯敏.用户社交网络平台信息分享行为研究－以上海市大学生微博使用为例[J].中国报业,2020(10):16－17.
［29］JIN L, LONG X, JOSHI J B D. Towards understanding residential privacy by analyzing users' activities in foursquare[C]// Proceedings of the 2012 ACM workshop on building analysis datasets and gathering experience returns for security. New York: ACM Press, 2012.
［30］杨世宏,陈堂发,张睿.隐私风险估算框架下用户位置信息敏感性认知[J].兰州大学学报(社会科学版),2019,47(4):8－16.
［31］中国网信网.第47次《中国互联网络发展状况统计报告》[R/OL].［2021－03－03］. http://www.cac.gov.cn/2021-02/03/c_1613923423079314.htm.
［32］TSAI J Y, KELLEY P, DRIELSMA P, et al. Who's viewed you? The impact of feedback in a mobile location-sharing application [C]// Proceedings of the SIGCHI conference on human factors in computing systems. New York: ACM Press, 2009.
［33］SADEH N, HONG J, CRANOR L, et al. Understanding and capturing people's privacy policies in a mobile social networking application[J]. Personal and ubiquitous computing, 2009, 13(6): 401－412.
［34］新浪财经.微博发布2020年第三季度财报[R/OL].［2021－03－03］. https://finance.sina.com.cn/stock/usstock/c/2020-12-28/doc-iiznezxs9422515.shtml.
［35］新浪微博数据中心.2018年微博用户发展报告[EB/OL].［2021－03－03］. https://data.weibo.com/report/reportDetail.
［36］中华人民共和国民政部.2020年中华人民共和国行政区划代码[EB/OL].［2021－03－03］. http://www.mca.gov.cn/article/sj/xzqh/2020/.
［37］STRANO M M. User descriptions and interpretations of self-presentation through facebook profile images[J]. Journal of psychosocial research on cyberspace, 2008, 2(2): 1－10.
［38］臧国全,周晓倩.隐私关注困扰下用户持续使用社交网络的动机解析[J].情报理论与实践,2020,43(3):71－77.
［39］杨嫚,温秀妍.隐私保护意愿的中介效应:隐私关注、隐私保护自我效能感与精准广告回避[J].新闻界,2020(7):41－52.
［40］卢家银,白洁.中国青年的网络隐私忧虑及其影响因素研究——基于对1599名共青团员的实证调查[J].新闻记者,2021(2):69－79.
［41］吴标兵,许和隆,张宇.中国公众隐私敏感度实证研究[J].南京邮电大学学报(社会科学版),2015,17(3):82－90.
［42］PREOTIUC-PIETRO D, COHN T. Mining user behaviors: a study of check-in patterns in location based social networks[C]// Proceedings of the 5th annual ACM web science conference. New York: ACM press, 2013.
［43］FIRE M, KAGAN D ELISHAR A, et al. Social privacy protector - protecting users' privacy in social networks[C]// The second international conference on social eco-informatics. Venice: IARIA XPS press, 2012.

作者贡献说明：

臧国全:提出研究思路,设计研究方案,修订终稿;
孔小换:图表绘制,修订初稿;
张凯亮:数据分析,撰写初稿;
于政杰:数据获取,数据处理。

Research on Social Network Users' Willingness to Self-Disclosure ——A Case of Sina Microblog

Zang Guoquan[1,2]　Kong Xiaohuan[1]　Zhang Kailiang[3]　Yu Zhengjie[1]
[1] School of Information Management, Zhengzhou University, Zhengzhou 450001
[2] Research Institute of Data Science, Zhengzhou City, Zhengzhou 450001
[3] School of Politics and Public Administration, Zhengzhou University, Zhengzhou 450001

Abstract: [**Purpose/significance**] Users' self-disclosure is of strategic significance to social network platforms based on user-generated content, and the quantity and quality of user-generated content depend on the user's willingness to self-disclosure. Therefore, the study of users' willingness to self-disclosure and its influencing factors can provide reference for social network platforms to formulate privacy policies and encourage users to disclose personal information, to promote the development of social networks platforms. [**Method/process**] Based on the existing research framework, a research model of social network users' self-disclosure willingness was constructed. This study took Sina microblog as an example, and adopted python crawler method to obtain users' personal data to analyzed users' willingness to self-disclosure. [**Result/conclusion**] Semantic content, location tags and data permission of microblog all affect users' willingness to self-disclosure. Hiding location tags and setting data permission can significantly improve users' willingness to self-disclosure. Social network users' willingness to self-disclosure is a kind of personal characteristics, which is affected by demographic factors, such as gender, age and education background of users.

Keywords: social network　self-disclosure　disclosure willingness　Sina microblog